Journal of Theoretical and Applied Information Technology
10th November 2015. Vol.81. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

GEOMETRY ALGORITHM ON SKELETON IMAGE BASED SEMAPHORE GESTURE RECOGNITION

[1]AERI RACHMAD, [2]MUHAMMAD FUAD
[1,2]Faculty of Engineering, University of Trunojoyo Madura, Indonesia
E-mail: [1]aery_r@yahoo.com

ABSTRACT

Semaphore, a way of communicating remotely, usually practiced in scouting activities. Information is delivered by gestures or movements using specific tools such as flags, paddles or rods. Teacher and instructors are needed for learning semaphore in conventional way as they will give examples and make correction when such an error occured. Based on the practical need to provide an alternative way to learn semaphore, this research proposes the use of geometry algorithm to develop a semaphore gesture recognition based on skeleton images that read from Kinect sensor. Euclidean distance and law cosines are two formulas that applied to generate gesture parameters of each alphabet. Recognition is achieved by comparing a pair of values of model and real-time gesture. Accuracy of this system that have been measured using RMSE with 30° of tolerance yields 90.76% for Alphabet and 88% for Word.

Keywords: *Semaphore Gesture, Geometry Algorithm, Kinect, Skeleton Image, Alphabet*

1. INTRODUCTION

There is a way for people to communicate visual distance by using two flags which is called as semaphore code. Mariners use this way of communication to transmit information from their ship to other mariners in the other ship [1]. Mariners like to use it because it is very useful give illumination during the night while during the day it is also used [2]. Scouts usually use semaphore in their activities by learning its flag with the correct angle [3].

By using semaphore, people may send messages, information or news with a special gesture using the gesture-flag, paddle, or rods. Scouts regularly use of a flag with size 40 x 40 cm attached to a stick measuring 50 cm. Flag color is hoisted red and yellow cross [4]. People can learn how to use Semaphore generally throuh the guidance of a teacher or a instructor that gives example and corrections when an error occurs in the gesture. Figure 1 shows a list of gestures that represent each letter in the alphabet.

Semaphore gestures have been developed in the form of a computer program [6]. It results semaphore gestures images which represent letters. However, this system still has weakness in that it [6] cannot represent the role of instructor for guiding the users to use semaphore. In other words, there is no interactive ability for user to interact with system.

The previous study of Noriyuki Iwane which was themed by the movement of the arm to detect katakana utilized kinect sensor and Open NI / NITE library.[13].

The weaknesses of this study [13] are: (1) There is still no rules about symbols in Japanese letters, (2) There is still the error or no detection of hand movements because there is no pause time between the characters to one another which results in confusion in the provision results, (3) There is a little pose variation of user in doing movement variation.

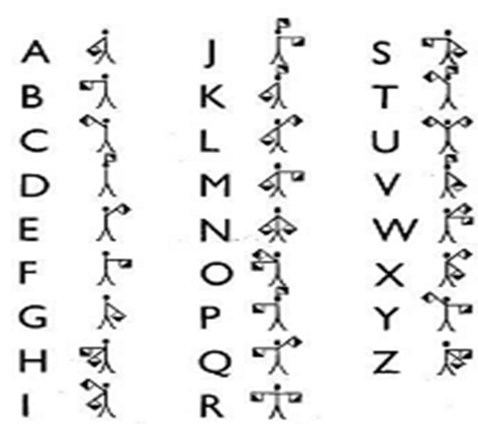

Figure 1. Gesture of each alphabet [5]

In this research, we give time limitation for 5 seconds to know every alphabet, we also uses Microsoft SDK library design with the help of C# programming language. In this research there has

Journal of Theoretical and Applied Information Technology
10th November 2015. Vol.81. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

been movement guidance from A to Z and also words variation so the users can have movement variations.

This research proposes the development of the ability to recognize gestures semaphore by utilizing the Kinect sensor. First of all, user stands in front of the sensor that is connected to the computer. Then, kinect sensor captures gestures of the user and presents it in the form of skeleton image stream. This image is composed of bone and joint. After that, semaphore gesture recognition program reads gestures parameters of image skeleton and match with the values that have been obtained from the model of semaphore gestures. Gesture recognition ability utilizes geometry algorithm to process variables consist of length of the bone and joint angles.

Kinect sensor is available in the gaming market by its official name, Kinect Sensor for Xbox 360. This interactive game controller was made by Microsoft that was built specifically to be connected with the XBOX 360 game engine. Figure 2 shows the components of the Kinect sensor that consists of a IR Emitter, LED, Color Camera, IR Depth Sensor, Microphone Array, and motors.

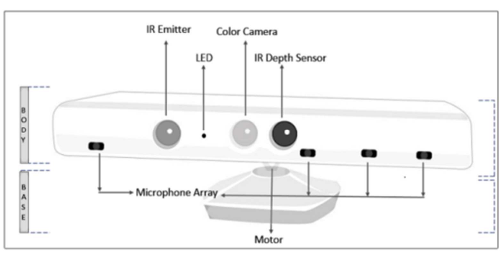

Figure 2.Components of the Sensor Kinect [7].

Table1 shows the specifications of the sensor that consist of the angle in the horizontal and vertical viewing, the number of frames that is generated per second, the number of dots in the image of the skeleton as well as the range of distance that is required in the process of tracking skeleton.

Table 1. Specification of Kinect Sensor [8]

Component	Spesifications
Viewing Angle	43˚ vertical 57˚ horizontal
Frame Rate	30 frames per second
Color Image resolution	640 x 480 pixel
Depth Image resolution	320 x 240 pixel
Skeleton image	20 points
Skeletal Tracking	1.2 to 3.5 meters

Kinect sensor produces three images are skeleton, Depth, and RGB, as shown in Figure 3 from left to right.

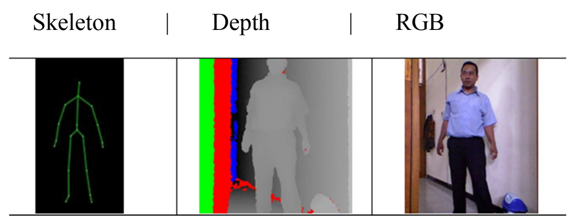

Figure 3.Image data acquisition

This sensor has been used to identify five gestures in the control of a wheeled robot to move forward, backward, stop, turn right, and turn left [9].

Wheeled robot motion control with the type of RGB-D camera is used to read the gestures and voice commands [10]. Interaction which is based on natural body language between humans and humanoid social robots using the same sensor was investigated in the previous research [11]. Utilization of Kinect to detect hand gestures using Depth image had also been investigated in the previous research [12]. Three of those four studies have used skeleton image in the gesture recognition process. In contrast to previous studies, this research utilizes skeleton image to detect semaphore gestures.

The image of the skeleton is a representation of the position and orientation of the player at any time.There are 20 points in skeleton image that represents the joint position of the human body. The 15 points that are used in data acquisition process in this research is shown in Figure 4.

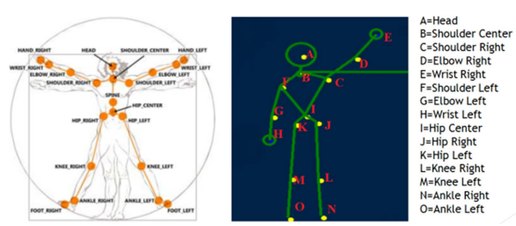

Figure 4.The position of the joint in the skeleton image[4].

Based on the correspondence between the semaphore gesture model and joint arrangement in the skeleton image, as shown in Figure5, this study uses only three main points. Center point of shoulder, right wrist, and left wrist are in accordance with the three points in the model gestures, i.e. the position of the center of the

Journal of Theoretical and Applied Information Technology
10th November 2015. Vol.81. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

shoulder, the flag in his right hand, and a flag in the left hand. The three main points can be constructed as a triangle.

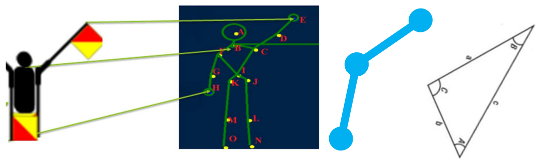

Figure 5.Correspondence between semaphore gesture model and joint arrangment to form a triangle [4].

Thus, the angle between the right wrist joint and left wrist which is measured relative to the joint center shoulder can be calculated by using geometry algorithms.

The organization of this paper is divided into some parts. Design of the system is described in section two. Section three consists of experiments and results. While conclusion is presented in section four. In section five, it consists of suggestions for further research development.

2. SYSTEM DESIGN

Two methods of geometry are used in this study, they are namely bone length measurement, and the measurement of the angle between the joint. Bone length that states the distance of joint skeleton can be calculated using Equation (1).

$$d_{2D} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

d_{2D} notation which is considered as bone length expressed the distance between joint in two-dimensional plane. The variables x_1, x_2, y_1, y_2 contain the value of the coordinates of each joint pair that are connected to form bone. Two bones are arranged as the sides of a triangle that form the angles between joint which the amount can be calculated with Equation (2).

$$\alpha = arccos\left[\frac{a^2 + b^2 - c^2}{2ab}\right] \quad (2)$$

α states the angle between two joints. a represents the side length of the triangle that faces angle α. While b and c respectively represent the length of two sides of the triangle that flanks α.

Root Mean Square Error (RMSE), which is frequently used to measure the difference between the value predicted by the model and the actual observed values of the environment that is being modeled is used to measure accuracy of the system. These individual differences may create residuals that serve to a single size aggregated- value in a predictive power [11].

The predicted model in the estimated X variable which is defined as the square root of the mean squared error formulates [13] as written in Equation (3):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}} \quad (3)$$

X_{obs} notation which considered as the value obtained from the results of the experiment. X_{model} is set value of the model.
The variable N contains Number of trials. RMSE values can be used to distinguish the performance of the model in a trial period with the level of validation as well as to compare the performance of individual models for the prediction model more [13].

Semaphore gesture recognition system by using the skeleton image requires reference data in the form of gestures parameters of the model. Figure semaphore as shown in Figure 1 is used as a model. Three gesture parameters, such as illustrated at Figure 6, that consist of α_L left hand corner (The angle formed by the 360˚ substract with left hand corner of the triangle as known as explementary angle), right hand corner α_R (The angle of the triangle formed by the right hand), and the difference between angle α_R and α_L.

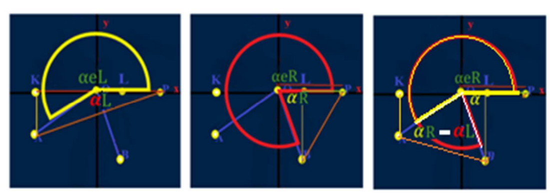

Figure 6. Process to measure gesture parameters.

The following will be exemplified how to get the parameters gestures by using the gesture model of the alphabet A. These parameters are obtained by applying geometry algorithm to each model of gestures. The process starts with determining the three main points of each image on the visual model of gestures. The three major points in each image are right wrist, left wrist and shoulder joint center, they are connected by line segments. Length of line segment that connects between the joint is called the bone length. This value is obtained by

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

calculating the root of the sum of squares of the difference between a pair of x and y coordinates of the second joint. The way to get the length of this bone, as is written in Equation (1), which is known as the Pythagorean Theorem. The second segment that joins these three main points form the angle of the Law of Cosine that can be calculated using Equation (2). For example as Figure 7: (a.) shows how to get a bone; (b.) shows how to get the angle α_L of the triangle AOP; (c.) shows how to get the angle α_L of the triangle BOP; (d.) demonstrates how to measure the angle between the triangular . α_R and α_L to get the triangle AOB. The results of measurements of parameters such as the amount gesture form angle in degrees. The three parameters consists of right hand corner α_R, α_L left hand, and the difference between α_R and α_L. These parameters are stored as a pattern as shown in Table 2.

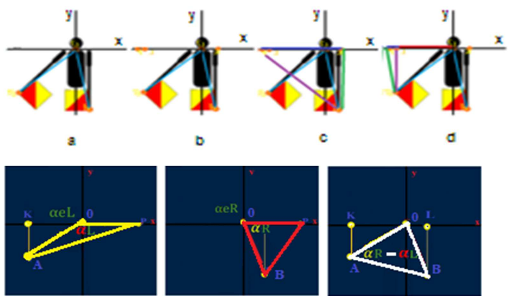

Figure 7.The process of getting the parameters of the model gesture.

3. EXPERIMENTS AND RESULTS

Based on the reference value of each gesture alphabet letters as listed in Table 2, the test are conducted to the gesture that is performed by users. Parameters gestures of the user skeleton image is compared with the parameters of the model. Process flow of gesture recognition based on geometry algorithms using skeleton image in real-time semaphore gesture application is shown in Figure 8.

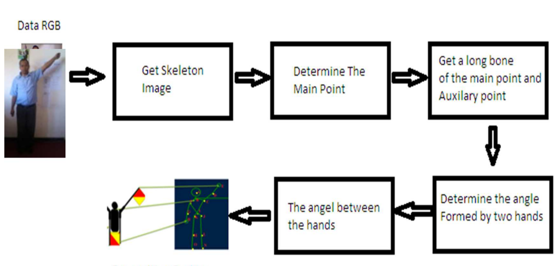

Figure 8. Real-time semaphore gesture recognition.

Table 2. The magnitude of the angle of gestures [4].

Alphabet	Angle (degree)		
	α_R	α_L	α_R - α_L
A	320.73	247.60	73.13
B	350.64	250.52	100.12
C	32.44	250.08	217.64
D	62.51	249.02	186.50
E	291.17	149.35	149.35
F	289.01	192.17	96.85
G	294.97	224.03	70.94
H	353.25	294.21	59.04
I	331.93	49.27	275.00
J	60.52	187.99	127.47
K	322.63	109.41	322.63
L	314.28	142.79	171.49
M	317.29	186.65	130.64
N	319.36	225.00	94.36
O	355.01	52.31	302.70
P	357.40	117.70	239.70
Q	351.72	143.68	208.04
R	352.83	190.95	161.88
S	355.03	232.32	122.71
T	35.95	99.64	63.69
U	29.64	155.19	125.55
V	67.26	224.03	156.77
W	134.62	182.34	47.72
X	139.57	219.24	79.66
Y	41.57	188.89	147.33
Z	243.87	185.81	58.06

Tests on the users were done in a room that has been laid out according to specifications of the sensor as in Table1. Indoor setup sensor test is illustrated in Figure 9. Laptop 2.3 GHz Intel CoreI 32 GB RAM equipped with an IDE Visual Studio 2010 and Kinect SDK is used in this trial. Sensor is placed at 1.5 m height from ground level. Users of the system stand by facing the sensor with a distance of 2 to 2.5 m. The room which is used in the test has a size of 3x4 m^2.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

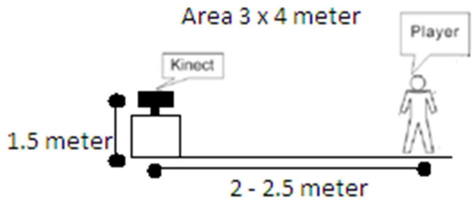

Figure 9. The arrangement of the sensors in the test room

Each user demonstrates his gestures of each alphabet letter. The value of the speed of the system to recognize gestures semaphore for real-time alphabet letters is presented in units of seconds. There are 21 letters that are recognized less than 5 seconds, 4 letters identified exactly 5 seconds, and 1 letter recognizable more than 5 seconds. RMSE values that are obtained from the experiment result average error values of each tolerance (These results are obtained from the average value of each error tolerance of 0 °, 10 °, 20 ° and 30). Results of the RMSE calculations are displayed in Table 3.

Table 3. RMSE of experiments.

Alphabet	Error RMSE							
	0°		10°		20°		30°	
	R	L	R	L	R	L	R	L
A	6.6	2.4	4.4	1.1	0	0	0	0
B	5.8	2.1	1.5	0	0	0	0	0
C	10	7.8	6.0	0	4.3	0	0	0
D	5	2.1	1.6	0	4.40	0	0	0
E	4.2	8.1	0	5	0	2.3	0	0
F	7.5	13	5.0	5.4	0	1.9	0	0
G	10.	5.8	0	3.2	0	2.03	0	0
H	2.2	3.1	0	6.3	0	3.1	0	0.9
I	9.7	12	7.2	7.9	6.6	3.1	0	0.05
J	5.7	9.3	6.1	4.14	3.3	1.9	0	0
K	9.7	6.9	6.5	4.08	3.5	1.8	0.9	0.39
L	6.2	7.2	8.6	4.42	5.7	1.8	1.1	0.04
M	11	7.7	5.5	3.9	4.4	0	0	0
N	6.6	6.6	3.2	4.12	0.9	1.8	0	0
O	3.8	8.9	0	4.23	0	4.2	1.1	3.3
P	5.2	7.6	0	3.42	0	1.83	0	0
Q	10.5	2.7	0.8	2.77	0.49	0	0	0.9
R	2.8	4.8	0	1.58	0	0	0	0
S	2.1	5.1	2.1	2.32	0	2.32	1.83	0.2
T	9.2	7.1	5.5	1.39	2.4	1.4	2.1	3.2
U	13	6.3	2.9	0	0.54	0	0	0
V	7.3	8.6	4.2	2.8	3.85	2.8	0	1.45
W	8.1	12	4.7	0.73	2.8	0.73	0	2.4
X	6.2	7.9	5.5	3.8	4.64	3.8	1.8	0
Y	10	4.2	7.2	0	4.5	0	0	0
Z	8.8	10	5.6	4.31	3.1	3.8	1.7	2.1
Average	7.88	7.64	4.54	3.64	3.04	2.1	0.82	1.16

The smallest average error, 0.82, value lies in the 30° tolerance for the right hand and 1.16 for the left hand. While 20° tolerance has average error 3.04 to 2.1 for the left hand and left hand. At 10° tolerance values obtained greater error is 4.54 for the right hand and 3.64 for the left hand. And more recently in the position of 0° indicates the greatest value error value is 7.88 to 7.64 for the left hand and right hand.

From the data above, the experimental results which were obtained by the average accuracy are the largest application at angles of 30° range and 0° low for all levels in alphabet and words; it can be seen in Table 4.

Tabel 4. The Average Accuracy of Letters and Words Prediction Level

Test	Tolerance			
	30°	20°	10°	0°
Letters	90.76%	72.30%	52.30%	23.07%
Words	88%	64%	48%	15%

4. CONCLUSION

Based on the results test, it shows that the semaphore gesture by utilizing the parameters which is stored in the skeleton image of the Kinect sensor reading scan be identified by using geometry algorithms. The results show significant Accuracy of Word Prediction using RMSE with 30° of tolerance yields 90.76% for Alphabet and 88% for Word.

But in this study there are still weaknesses as time for the introduction of the alphabet is still too long so that the user can still move the arm in order to recognize letters alphabet. Besides the difference between katakana and alphabet Indonesia, which makes many variations of movement or pose. The amount of tolerance angle that is too small causes the error committed by either the left hand or the right hand.

5. SUGGESTION

This research will be continued to recognize phrases that will be applied in the form of games using Semaphore. In addition to the field of multimedia, this study can also be developed to produce human computer interaction with capabilities base on Kinect sensor

REFERENCES:

[1]. Madore, A..2007. *Semaphore Notes, Canada's Naval History, Canadian War Museum.* diambil pada Senin, 16 Maret 2015 dari situs http://www.warmuseum.ca/cwm/exhibitions/navy/home-e.aspx.

[2]. Signal Corps United States Army 1910.2013.*Visual Signaling* (Ebook #43515). War Department Office of The Chief Signal Officer.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[3]. Seton, E. T. Chief Scout.2012. *Boy Scouts Handbook, the first edition 1911*. Chapter IV. Tracks, Trailing and Signaling. In William D. Murray, George D. Pratt, A. A. Jameson (Ed. Board).New York: Dover Publications, Inc.

[4]. Fuad, M (2015). Semaphore gesture detection using Kinect sensor. Seminar Nasional Aplikasi dan Pengembangan Teknologi Informasi.

[5]. Horizon R. H. 1958 (2006). Signalling for the Patrol. Dennis Trimble. Publish by The boy Scots Association.

[6]. Helena Sinaga (2015), Learning Application on Signal Scout Semaphore Multimedia And Web-Based Using Computer Assisted Instruction Method.) Jurnal Informasi dan Teknologi Ilmiah (INTI) Volume : V, Nomor : 3 , Pp 160-164

[7]. Abijit Jana. (2012). Kinect for Windows SDK Programming Guide. Birmingham: Packt Publishing.

[8]. Microsoft Research. (2011). Kinect for Windows SDK Beta Programming Guide.

[9]. Broccia, G., Livesu, M., Scateni, R. (2011). Gestural Interaction for Robot Motion Control. Eurographics Italian Chapter Conference. Eurographics Association.

[10]. Fardana, A.R., Jain, S., Jovancevic, I., Suri, Y., Morand, C., Robertson, N.M. (2013). Controlling a Mobile Robot with Natural Commands based on Voice and Gesture.

[11]. Zhang, Z., Xiao, Y., Beck, A., Yuan, J., Thalmann, D. (2014). Human-Robot Interaction by Understanding Upper Body Gestures.

[12]. Fuad, M. (2014). Development of Hand Gestures Based Detection Using Depth Image Matching Features. Seminar Nasional Ilmu Komputer. Universitas Gadjah Mada.

[13]. Noriyuki Iwane (2012). Arm Movement Recognition for Flag Signaling with Kinect Sensor. IEEE International Conference Virtual Environments human Computer Interface and Measurement Systems (VECIMS). page 86 - 90